A scientometric exploration of crowdsourcing: Research clusters and applications

Sercan Ozcan [a, b, c], David Boye [a], Jbid Arsenyan [d], Paul Trott [a]

[a] Portsmouth Business School, University of Portsmouth, Portsmouth, UK

[b] Department of Engineering Management, Bahcesehir Universitesi, Istanbul, Turkey

[c] National Research University Higher School of Economics (Russian Federation), Moscow, Russia

[d] Rennes School of Business, Rennes, France

e-mails: {sercan.ozcan@port.ac.uk; david.boye@port.ac.uk; jbid.arsenyan@rennes-sb.com;

paul.trott@port.ac.uk}

A*bstract*

Crowdsourcing is a multidisciplinary research area, it represents a rapidly expanding field where new applications are constantly emerging. Research in this area has investigated its use for citizen science in data gathering for research and crowdsourcing for industrial innovation. Previous studies have reviewed and categorised crowdsourcing research using qualitative methods. This has led to the limited coverage of the entire field, using smaller discrete parts of the literature and mostly reviewing the industrial aspects of crowdsourcing. This study uses a scientometric analysis of 7,059 publications over the period 2006 - 2019 to map crowdsourcing research to identify clusters and applications. Our results are the first in the literature to map crowdsourcing research holistically. We classify its usage in the three domains of innovation, engineering, and science where 11 categories and 26 sub-categories are further developed. The results of this study reveal that the most active scientific clusters where crowdsourcing is used are Environmental Sciences and Ecology. For the engineering domain, it is Computer Science, Telecommunication and Operations Research. In innovation, idea crowdsourcing, crowdfunding and crowd creation are the most frequent areas. The findings of this study map crowdsourcing usage across different fields and illustrate emerging crowdsourcing applications.

Keywords: *Crowdsourcing, Scientometric, Text Mining, Emerging Clusters, Research trends*

I. INTRODUCTION

Crowdsourcing continues to be a growing area for researchers as an emerging phenomenon [1], defined as *the solving of a task or problem by a crowd of people outside an organisation using online interaction* [2]. It is the act of an organization taking a project or task, once performed by employees, and outsourcing it to an undefined (and generally large) network of people in the form of an open call [3]. It harnesses the power of the crowd in tasks such as idea creation and solving problems, and it does not necessarily require contracts of agreement found in outsourcing tasks to a specialised organisation [4, 5]. Internet-based platforms have provided the avenue to reach a far wider group of contributors that could result in better solutions [6]. These platforms have also led to a wide dispersion of crowdsourcing in different areas and so it is difficult to estimate the coverage of crowdsourcing applications.

Sivula and Kantola [7] divide crowdsourcing models into seven main categories: crowd-voting, idea crowdsourcing, crowd evaluation, crowd creation, micro-tasking, solution crowdsourcing and crowdfunding. Palacios et al. [8] propose its emerging research trends range from problem-solving, learning paradigms, open innovation program, new product development (NPD) and collaborative initiation. Hossain and Kauranen [9] use a literature review method based on 52 highly cited studies and they categorised the application areas as *idea generation, microtasking, open-source software, public participation, citizen science, citizen journalism,* and *wikis*. Zhao and Zhu [1] categorise the field based on the *focus* from 55 studies: conceptualization focus, the system focus, and the application focus.

Whereas these previous studies identify crowdsourcing research [1, 7-10], they lack a holistic approach where all crowdsourcing scientific domains are mapped and categorised. We have identified only one quantitative study on the topic [7] with a dataset of 346 studies, research being specific to the innovation management field. This emerging field lacks a holistic quantitative examination to map the entire field without domain-specific limitations. Such a study would reveal applications across different domains and intersections between them.

In terms of quantitative approaches, scientometric studies have been used in other fields to understand trends and the growth of scientific fields. Such a study remains absent within crowdsourcing. We also did not find any crowdsourcing specific data collection method or taxonomy. Based on the extensive literature review, we identify methodological and practical gaps in this field.

Those gaps in the literature form the basis of our research question: What are the crowdsourcing research fields? We aim to identify crowdsourcing clusters and categories to answer this question. Furthermore, we aim to illustrate knowledge management and benchmarking to related research fields as it identifies themes within them, provides the visualization of research data [11] and finally, offers trend and innovation analysis [12]. A scientometric approach is chosen for this study as it helps to identify the foundations of a research field by using a quantitative approach, minimizing the subjectivity of the results [13]. Accordingly, our study integrates the combination of scientometric and text mining methods to examine the structure and growth of crowdsourcing. We provide a search string and a method for other scholars to enable them to follow similar studies in crowdsourcing field.

Our findings contribute to the stream of literature on crowdsourcing by providing a scientometric-based methodological analysis of its use in the domains of science, engineering and innovation [1, 7-10]. Considering the relevant studies [1, 7-10], this is the only study in the literature that maps entire crowdsourcing research with a scientometric approach and it helps breaking down this area into its main and sub-domains. As a result, there are three main domains are found and these domains are further broken down to 11 sub-categories based on crowdsourcing usages, such as idea and wisdom in innovation, mapping in engineering, and habitat monitoring in science. Our findings also contributes to the literature by linking crowdsourcing applications to their relevant platforms and techniques. Our approach gauges more accurately the crowdsourcing field in term of uncovering crowdsourcing application typologies, classifying emerging and mature applications as well as clustering emerging research clusters according to term linkage strength within research fields, all with various implications for researchers, analysts and innovation specialists. Finally, our study contributes to the future research directions in crowdsourcing considering three main domains and its sub-categories.

The next section reviews the literature on crowdsourcing. The third section presents the analytical framework, followed by the research methodology in the fourth section. The fifth section presents our findings and analysis. The paper concludes with a few remarks.

II. BACKGROUND AND RELEVANT RESEARCH

Existing qualitative reviews of crowdsourcing research provide useful insights regarding definitions, models, applications and avenues for future research. Table I provides an overview of recent studies that advance our understanding of crowdsourcing [14], categorising the relevant crowdsourcing literature in three main categories as 1) examination of crowdsourcing characteristics, 2) utilization of crowdsourcing within research fields, and 3) taxonomy of crowdsourcing research. We illustrate different levels of crowdsourcing focus, starting from crowdsourcing definition level studies to more advanced use of crowdsourcing in various conditions and fields; ranging from examining the key determinant level for its implementation [1] and key components in a crowdsourcing process and systems [22], to crowdsourcing typologies and classifications [24, 27].

Table I clearly shows that the diversity of typologies and perspectives makes it difficult to provide a generic categorisation to crowdsourcing research and its applications. Furthermore, lack of a review based on word metrics makes it difficult for scholars to accurately examine the growth and development of crowdsourcing research. Thus, there is a need to better understand the boundaries of crowdsourcing research, as well as to identify quantify trends, metrics and to visualize fields and subfields.

Table I Categories of research on crowdsourcing

Focus	Research contributions	Authors
Examining of crowdsourcing characteristics	Definition of crowdsourcing Identifying the capabilities and tasks performed Identification of benefits of crowdsourcing on an individualistic and community level Identifying motivations for crowd involvement in crowdsourcing Identifying task characteristics and complexities (simple, creative and complex) Examining the perspectives of crowdsourcing (organisational, Technical, process, human-centric) Identifying components of crowdsourcing systems (User, Task, Contribution and Workflow management) Exploration of crowdsourcing typologies	[3], [15 - 19]
Utilization of crowdsourcing within research fields	The use of crowdsourcing in higher education (crowd wisdom, crowdfunding, crowd voting and crowd creation) Examination of crowdsourcing focus within information systems: Conceptualization, Application and System Pillars of crowdsourcing model in information systems The use of crowdsourcing in bioinformatics: tasks (Micro and Megatasks) and application systems (volunteer labour, purposive gaming, microtask markets and open innovation contests) Use of crowdsourcing in health and medicine: tasks performed (problem-solving, surveying, surveillance, monitoring and data processing) Understanding crowdsourcing in Human resource management: Jobs (routine, complex and creative tasks), workforce planning, training and development, recruitment fit (person-organisation, person-group and person-job), compensation, legal and ethics Use of crowdsourcing in agriculture: Tasks (knowledge, data and visual observations),	[20 - 27]
Taxonomy of crowdsourcing research	Understanding aspects of crowdsourcing: Application (voting system, information sharing, game and creative systems) algorithms, performance (user participation and quality management and cheating) and dataset Organization level (acceptance, implementation, management, quality, evaluation), technology level (incentive mechanisms, technological issues), and participation level (crowd motivation, organization employees' behaviours) Examining aspects of crowdsourcing: process, characteristics, motivation to participate, motivation to crowdsource and limitation Examining the models (intermediary, citizen media production, collaborative software development, digital goods sale, product design, p2p financing, consumer report, knowledge base and collaborative science project), issues (level of collaboration and type of service outsourced) and control mechanisms (compensation schemes, trust-building, voting and commenting) Breakdown of internal crowdsourcing (Problems, Governance and Outcomes) Breakdown of crowdsourcing process: Input (Problem/Task), Process (Session, People, Knowledge and Technology management), Output (Solution, Seekers and Solvers Benefits)	[1],[4],[9],[14], [28 - 32]

Analysis of the literature using objective metric tools and network analysis can provide new insights. The most common methods and techniques to review a field with a wide scope of studies are bibliometrics, scientometrics, and informetrics [33]. Recently, Malik et al. [32] examined crowdsourcing publications using parameters such as document type, language, prolific journals, leading countries, institutions and authors of publications. The merits and most significant reason for using this approach are to fathom the features of a scientific discipline. Scientometric analysis, on the other hand, is the second most used metric method for the analysis of past, present and future scientific developments of a field. This form of analysis is a quantitative method of science mapping used to analyse the existing intellectual core and landscape of a research field [34]. Such a study will reveal a pictorial trend of crowdsourcing research which will help scholars to understand how the field has developed over time. Accordingly, in the next section, relevant scientometric studies are examined to understand its merits and areas of application.

A. Relevant Scientometric Studies

Scientometric analysis can provide valuable information about the changes in trends, detect real-time hot topics and provide avenues for further research. This is achieved using data collection, visualization and science mapping interpretation tools on a field of study [34]. The dynamic visualization occurrence and citation technology enable the display of a knowledge domain to identify research areas and sub-areas by clustering on the map as a basis of knowledge.

Innovation literature proposes some interesting scientometric research. Su and Lee [35] propose a way of mapping the structure of open innovation research by analysing publications retrieved from the Web of Science. The study combines keyword co-occurrence with social network analysis to examine the country, institute, publication and keyword relationship; to identify that activation, connection dynamics, virtual communities, and open-source software are some of the dominating key terms in the field. Chatterjee and Sahasranamam [36] examine the trend in innovation management research in India during the period 1991 – 2013, to reveal the concentration of publications in general management/strategy (10%), entrepreneurship (5%) and innovation-related journals (85%) as well as the trending key themes focusing on macro perspectives, operational/technical aspects and organisation aspects. Appio et al [37] examine the main research areas of social media-based innovation to understand its development during the period 2003 – 2013. The study proposes emerging clusters within the field range from organisational learning, knowledge sharing in communities, value (co)creation, user/customer involvement in innovation process and, open and distributed innovation. Kullenberg and Kasperowski [38] performed a scientometric analysis on a particular research field within crowdsourcing which is known as citizen science, to explain its development over time and to show what strands of research have adopted citizen science. Mora et al. [39] examine the development path of smart cities in an attempt to visualize the network of publications shaping the structure of the field, revealing emerging paths and mapping the thematic clusters. Similarly, Kovács et al. [13] reveal seven thematic clusters within open innovation field: (1) firm-centric aspects of Open Innovation, (2) management of Open Innovation networks, and (3) role of users and communities in Open Innovation. This scientometric review of the field reveals the firm-centric aspects of Open Innovation focused on the role of knowledge, technology, and R&D from the innovating firm's perspective whilst the management of Open Innovation networks and role of users and communities in Open Innovation had been relatively under-researched.

These studies show that scientometric analysis needs to be combined with an analysis of recent trends in academic research. Given the gaps identified earlier, the aim of this study is to examine and map where crowdsourcing research and applications are concerned, specifically with regards to knowledge domains, sub-domains and its relevant industries. Hence, the following research objectives emerge as 1) to identify crowdsourcing main clusters and visualise the network of research, 2) to study the link between research cluster and sub-clusters, 3) to examine crowdsourcing applications within clusters, and 4) to build a framework that illustrates the emerging clusters and sub-clusters within the crowdsourcing field. Figure 1 demonstrates the process followed to assist in achieving the objectives of this study.

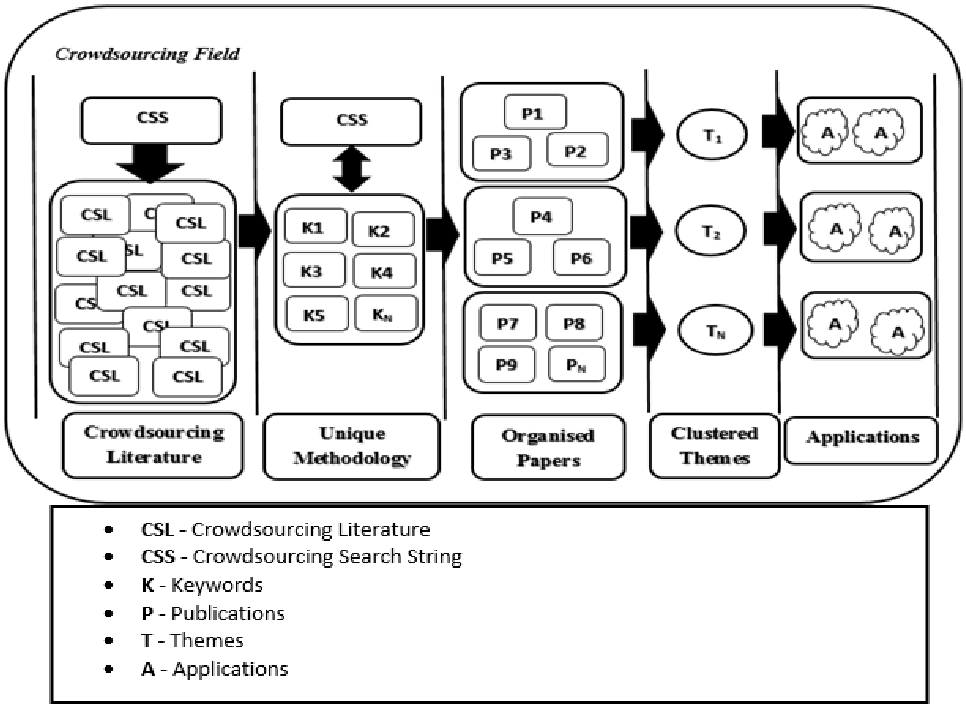

Fig 1: Conceptual framework for deriving themes and crowdsourcing applications [11 - 13]

III. RESEARCH METHODOLOGY

Considering the methodological gap of this literature and following the key approaches [87-88], this study uses scientometrics to examine crowdsourcing literature. Examination of other similar methodological approaches show that there are four or five similar steps starting from data collection to data interpretation steps [87-90]. Considering these studies, our methodological approach is designed. Figure 2 illustrates the process followed for this study where as a result, the entire crowdsourcing research is mapped and future directions are provided. Accordingly, this study has 5 major steps. In step 1, publication data is collected by identifying the key search terms, identifying inclusion and exclusion criteria and finally retrieving the data. In step 2, the collected publication data is optimised by cleaning irrelevant terms and allocating thresholds before terms are examined. In step 3, the matrix of terms are calculated based on co-occurrence and then positioned based on the centrality measures. In step 4, the centrality measures are illustrated based on the relevance scores and also the frequency of terms using clustering and heatmap visualisation interfaces. In step 5, the results are crosschecked with the database and also the search string is updated then we performed step 1-4 once again. In step 6, the results are interpreted by using combinations of terms in the final database to identify relevant studies and these are read through to understand the cluster of terms and also overlapping clusters with each other. The details of data collection, analysis and verification steps are explained in the following sections.

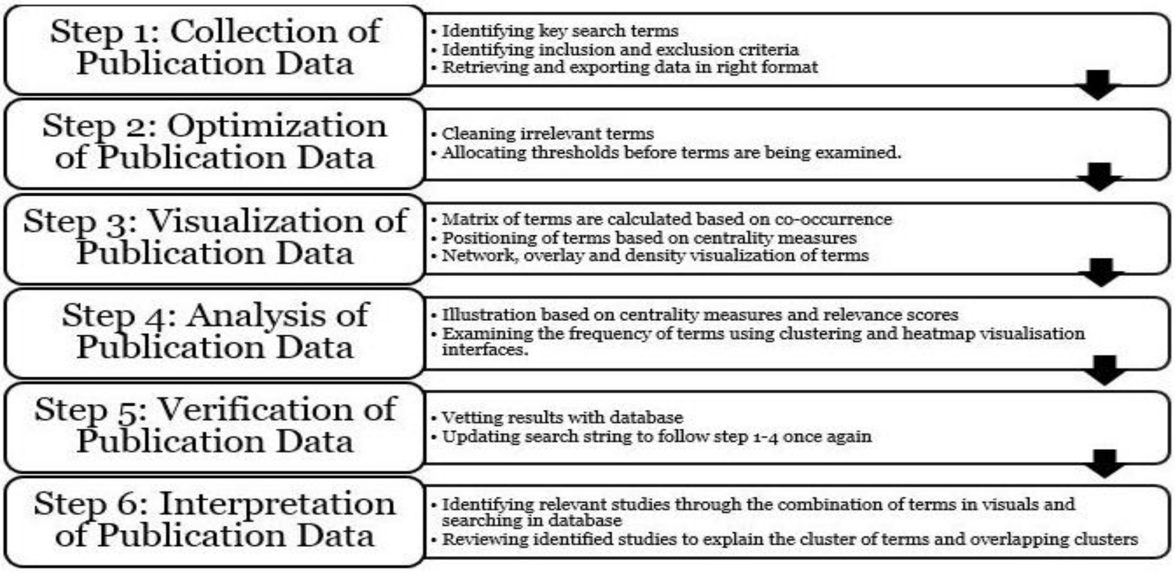

Fig 2: Methodology for publication data analysis

A. Data Collection and Verification

The Web of Science (WOS) database was utilised as the source of data for the period 2006 – 2019. As crowdsourcing is a multidisciplinary field with a variety of emerging applications, we decided to use a search string in the “topic” field of articles which allows the accumulation of title, abstract and keywords as they play an important role in the scientometric analysis. Initially, using the terms that are identified in the literature review, we gathered all the crowdsourcing specific terms. The preliminary analysis is completed to uncover new terms in the results to expand the search string. The search string is finalised after three iterations following the previously mentioned step 1 to 4 with the new crowdsourcing specific terms each time to reach to the saturation of crowdsourcing terms. The final search string is as shown below:

“crowdsourc OR crowd-sourc* OR “crowd sourc*” OR “mobile crowdsourc*” OR crowdfund* OR "crowd fund*" OR “crowd financ*” OR crowdfinanc* OR “crowd vot*” OR crowdvot* OR “crowd scienc*” OR citizenscienc* OR “citizen scienc*” OR “crowd test*” OR crowdtest* OR “crowd mapp*” OR crowdmapp* OR “crowd sens*” OR crowdsens* OR “crowd comput*” OR crowdcomput* OR “crowd solv*” OR crowdsolv* OR ((macrotask* OR “macro task*” OR “micro task*” OR microtask*) AND crowd)”*

An initial search for the period revealed 13,371 articles but after further processing and limiting documents to only English language scientific articles and eliminating conference proceedings, other review studies or book chapters, we arrive at a total of 7,059 articles for this area.

Based on step 3 and 4, after analysing the data, we identify three major clusters where crowdsourcing is applied: innovation, scientific research and engineering domains as shown in Table IV, the details of which are explained in the findings section in Figure 4. To examine these three identified major clusters in-depth and to break them down to their own subcategories, we define three separate search strings for each cluster, as shown in Table II. For each string, we make use of the crowdsourcing general query to have consistency across different clusters, then we differentiate each cluster with their dedicated subset of search strings. These cluster-specific terms are expanded using the iteration method (preliminary analysis are completed for each cluster with the initial terms then expanded with three iterations). Moreover, the search strings are further supported by WOS categories, broken down in Table II.

Table II: Breakdown of Search Strings

Clusters	Cluster Search String	Research Categories
C1 - Crowdsourcing and Innovation	crowdsourc* OR crowd-sourc* OR “crowd sourc*” OR crowdfund* OR "crowd fund*" OR “crowd financ*” OR crowdfinanc* OR “crowd vot*” OR crowdvot* OR “crowd test*” OR “crowdtest*” OR “crowd solv*” OR crowdsolv* OR ((macrotask* OR “macro task*” OR “micro task*” OR microtask*) AND crowd) NOT “crowd scienc*” NOT citizenscienc* NOT “citizen scienc*” NOT “mobile crowdsens*” NOT “crowd mapp*” NOT crowdmapp* NOT “crowd sens*” NOT crowdsens* NOT “crowd comput*” NOT crowdcomput*	Management, Business, Art, Operations Research and Management science, Economics, Health Care Science Services, Psychology Multidisciplinary, Law, Business Finance, Hospitality Leisure Sport and Tourism, Green Sustainable Science Technology, Ergonomics, Political Science, Engineering Manufacturing, Infectious Diseases, Public Administration, Food Science Technology, Social Sciences Biomedical, Engineering Industrial
C2 - Crowdsourcing and Engineering	crowdsourc* OR crowd-sourc* OR “crowd sourc*” OR “mobile crowdsourc*” OR “mobile crowdsens*” OR “crowd mapp*” OR crowdmapp* OR “crowd sens*” OR crowdsens* OR “crowd comput*” OR crowdcomput*OR ((macrotask* OR “macro task*” OR “micro task*” OR microtask*) AND crowd) NOT “crowd scienc*” NOT citizenscienc* NOT “citizen scienc*”	Computer Science Information Systems, Telecommunications, Engineering Electrical Electronic, Computer Science Artificial Intelligence, Computer Science Software Engineering, Computer Science Theory Methods, Computer Science Interdisciplinary Applications, Computer Science Hardware Architecture, Instruments Instrumentation, Transportation Science Technology, Chemistry Analytical, Engineering Civil, Regional Urban Planning, Computer Science Cybernetics, Electrochemistry, Engineering Environmental, Transportation, Automation Control Systems, Imaging Science Photographic Technology, Acoustics, Behavioural Sciences, Physics Applied, Mathematics Interdisciplinary Applications, Medical Informatics, Language Linguistics, Linguistics, Information Science Library Science, Materials Science Multidisciplinary, Construction Building Technology, Neurosciences, Psychology Experimental, Engineering Multidisciplinary, Radiology Nuclear
C3 - Crowdsourcing and Science	crowdsourc* OR crowd-sourc* OR “crowd sourc*” OR “crowd scienc*” OR citizenscienc* OR “citizen scienc*”	Ecology, Environmental Sciences, Biodiversity Conservation, Multidisciplinary Sciences, Communication, Geography, Environmental Studies, Remote Sensing, Geography Physical, Public Environmental Occupational Health, Geosciences Multidisciplinary, Marine Freshwater Biology, Zoology, Water Resources, Meteorology Atmospheric Sciences, Astronomy Astrophysics, Biology, Education Educational Research, Entomology, Ornithology, Urban Studies, Evolutionary Biology, Sociology, Oceanography, Plant Sciences, Mathematical Computational Biology, History Philosophy of Science, Fisheries, Forestry, Biochemistry Molecular Biology, International Relations, Chemistry Multidisciplinary, Genetics Heredity, Development Studies, Medicine General Internal, Biotechnology Applied Microbiology, Geochemistry Geophysics, Biochemical Research Methods, Statistics Probability, Education Scientific Disciplines, Psychiatry, Humanities Multidisciplinary, Psychology-Clinical, Pharmacology Pharmacy, Medicine Research Experimental, Substance Abuse, Surgery, Microbiology, Biophysics, Clinical Neurology

To increase the validity and the reliability of the results, we implement quantitative and qualitative examinations at different stages of the study. To increase the validity, the search string is expanded as much as possible with a number of iterations between data collection and data analysis as mentioned previously. Further, we restrict the usage of terms to be only specific to the crowdsourcing field. In other words, we do not use any terms that may lead to irrelevant studies. To increase the reliability, each time a new term is introduced, we list the results from the least relevant to the most relevant and examine to see the relevancy of the studies. For instance; the term "macro task" was leading to the inclusion of unnecessary studies, so we combine it with the term "crowd" to limit it to the crowdsourcing specific studies where "macro task" terms are used. 10% of the final set of data is reviewed using the reliability ranking and we had less than 1% error in the data. Even this minimal error is lowered with minimal thresholds and co-occurrence terms in the following sections. In a cluster-specific in-depth analysis, research categories increase the reliability and validity even further when three are clusters examined separately.

B. Data Analysis, Visualisation and Interpretation

We use the title and abstracts of the collected data to study the results. Using the terms in these sections, we create a co-occurrence of terms with a minimum threshold of at least 10 occurrences based on binary counting. Generic or irrelevant terms are eliminated by using stopwords and relevance scores. Afterwards, any terms without any co-occurrence are eliminated to minimise outliers and to provide a better representation of clusters based on the interlinkages between terms. As a result, a total of 297 unique keywords identified for further analysis.

We visualize and examine clustered domains of the crowdsourcing field using centrality measures with the frequency of terms. Using the VOSviewer software, we visualise the results using the clustering feature in the density visualisation. The results are normalised based on association strength. The clustering results are represented graphically using cluster density maps. Clustering numbers are organised by VOSviewer where small clusters are merged. These are displayed in a variety of ways such as the label, density, cluster density and scatter view. In the clustering visuals, the high term density represents high term frequency and hence a high number of research outputs in those research areas. The terms are represented using occurrence based weights and accordingly the individual font size of terms also represents the frequency of those terms across all the abstracts and titles of the studies. The closeness of two or multiple terms illustrates a high relationship of terms with each other due to common usage of them in abstracts and titles.

Consequently, we interpret our final visuals using an in-depth qualitative examination approach. To perform this step, we select the combination of terms that are illustrated by the clustering results and then searched the relevant articles in the database. We select a minimum of 5 articles for each interpretation, considering the citation score of the study as well. After reviewing each cluster, we label them accordingly. Finally, each cluster is categorised, and then these categories are linked with the relevant applications to illustrate the crowdsourcing practices across the three different clusters that are identified in our study.

IV. RESULTS AND FINDINGS

The results are organised from descriptive to more in-depth analytical findings. Firstly, the studied area is categorised and the top terms are illustrated. Secondly, the crowdsourcing area is clustered. Thirdly, the main clusters are examined separately and then categorised into sub-groups. Finally, the identified crowdsourcing categories are illustrated with their relevant applications.

A. Crowdsourcing Research Areas

Based on the WOS research areas, Figure 3 illustrates the productive fields with high contributions to crowdsourcing research during 2006 – 2019.

Computer science leads the crowdsourcing research by 2,035 publications making up 26% of all research during the period of observation. The reason for this significant contribution is twofold. Firstly, crowdsourced efforts are a major component in software development with platforms such as Github, FLUX and so on [40, 41]. Secondly, crowdsourcing is heavily linked with artificial intelligence research, a major domain in computer science which is increasing in reach as machine learning capabilities are expanded with the use of the crowd in data labelling [42] and the access to mobile devices for crowdsensing [43].

The leading field for crowdsourcing scientific research is Environmental Sciences and Ecology with 1,058 publications, 15% share of the dataset. This substantial share is easily explained by the rise of the involvement of citizen scientist in environmental research along with other environmental subfields such as biodiversity conservation (405 publications, 5%), Zoology (195 publications, 2%) and marine freshwater biology (158 publications, 2%).

When we proceed to a broader field: Engineering with 1,097 publications making up 14% of all research in the dataset. Even though the field represents the 3rd most represented field, it is mainly due to part of engineering research is overlapping with other fields mainly computer science (2,035 publications, 26%), telecommunication (674 publications, 9%), science and technology (427 publications, 5%) operations research and management science (75, publications, 1%) and transportation (101 publications, 1%).

Fourth place belongs to business economics with 696 publications, representing 10% of research during the observed period. This can be explained by considering the significance of crowdfunding as a major sub-cluster of crowdsourcing research. This influences fields such as management, business, and entrepreneurship. In addition, Innovation management and new product development are other fields which may explain why crowdsourcing applications are high in this research area.

The diversity of research fields in Figure 3 presents visual evidence that crowdsourcing research is multi-dimensional and interdisciplinary; it combines different perspectives, theories, and a variety of applications to solve complex problems.

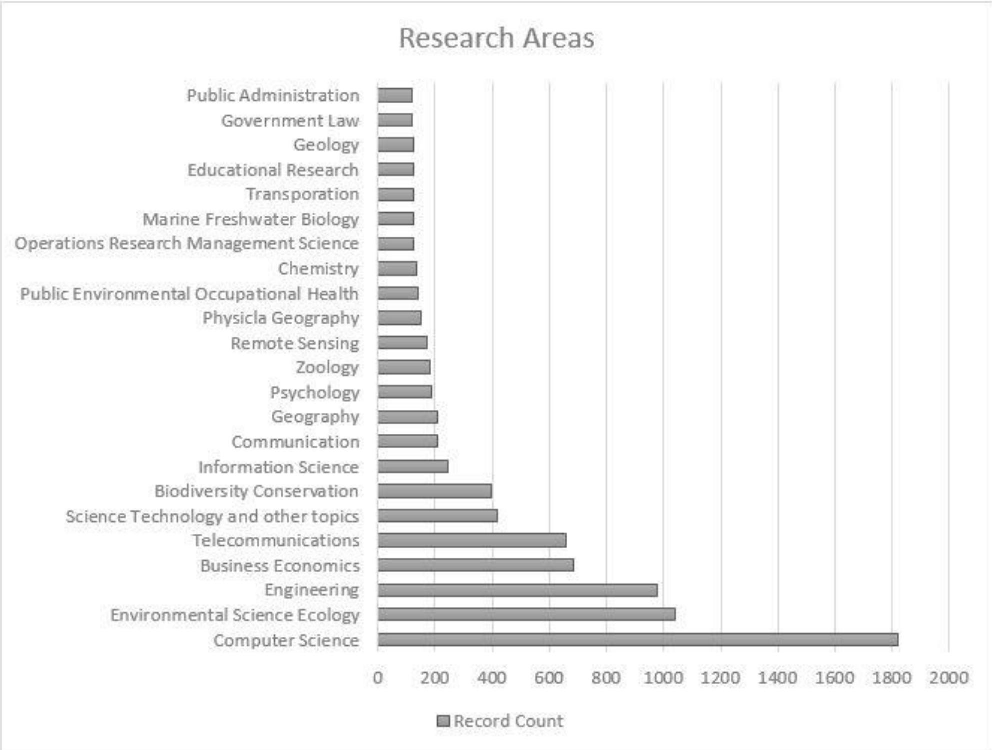

Fig 3: Breakdown of crowdsourcing scientific research areas

B. Keyword Analysis

Once we identified the major research fields in the dataset, we applied keyword analysis, in order to have a comprehensive overview of the overall theme, before identifying the clusters. A keyword analysis is an effective way to explore topical emphases and we use text mining methods to this effect. Terms were extracted by natural language processing. We then filtered the extracted terms to remove (a) common words used within our search string and (b) words with little to no relevance such as *research methodology, conclusion, future research,* *literature review,* etc. After this data cleaning process, a total of 297 unique keywords emerged. We distinguish the *Top 15 Keywords* for each cluster via keyword analysis, shown in Table III. These keywords, along with the cluster map in Figure 4 in the next section, enable the identification of the main research themes within the clusters. Cluster 1 (C1) illustrates keywords that are related to the application areas of crowdsourcing in the innovation field, Cluster 2 (C2) illustrates it for the engineering field and Cluster 3 (C3) illustrates it for the scientific research field (citizen science).

Table III: Keyword Statistics

	Cluster 1 - Crowdsourcing and Innovation		**Cluster 2 - Crowdsourcing and Engineering**		**Cluster 3 - Crowdsourcing and Science**	
	Terms	Occurrence	Terms	Occurrence	Terms	Occurrence
1	Idea	551	Task	1602	Volunteer	926
2	Concept	464	Algorithm	1306	Site	876
3	Campaign	462	Performance	1053	Observation	849
4	Social medium	423	Worker	862	Pattern	719
5	Product	382	Solution	746	Conservation	444
6	Motivation	377	Sensor	615	Detection	441
7	Reward	274	Device	445	Habitat	440
8	Firm	255	Smart device	405	Temperature	319
9	Contest	216	Classification	398	Biodiversity	304
10	Rating	207	Privacy	319	Ecosystem	262
11	Language	199	Mobile crowdsensing	226	Road	173
12	Entrepreneur	198	Machine	180	Surface	84
13	Regulation	149	Computation	170	Monarch	61
14	Fund	121	Mobile user	140	Water quality	50
15	Developer	91	Payment	90	Opportunistic data	39

Table III provides an overview of popular crowdsourcing terms for each cluster. For instance, for C1, it is apparent that crowdsourcing is more commonly used for the ideation process (‘idea’ is ranked as the top one in C1) in innovation and new product development compared to regulation oriented applications (‘regulation’ is ranked as the 13th most common word in C1). However, many are interrelated with each other and these interlinkages are investigated in the following sections.

C. Cluster Analysis

Co-occurrence map in Figure 4 displays quasi-connected clusters C1 and C2 and a relatively independent cluster C3 in crowdsourcing research. The next three sections discuss the general themes and justify these main crowdsourcing clusters.

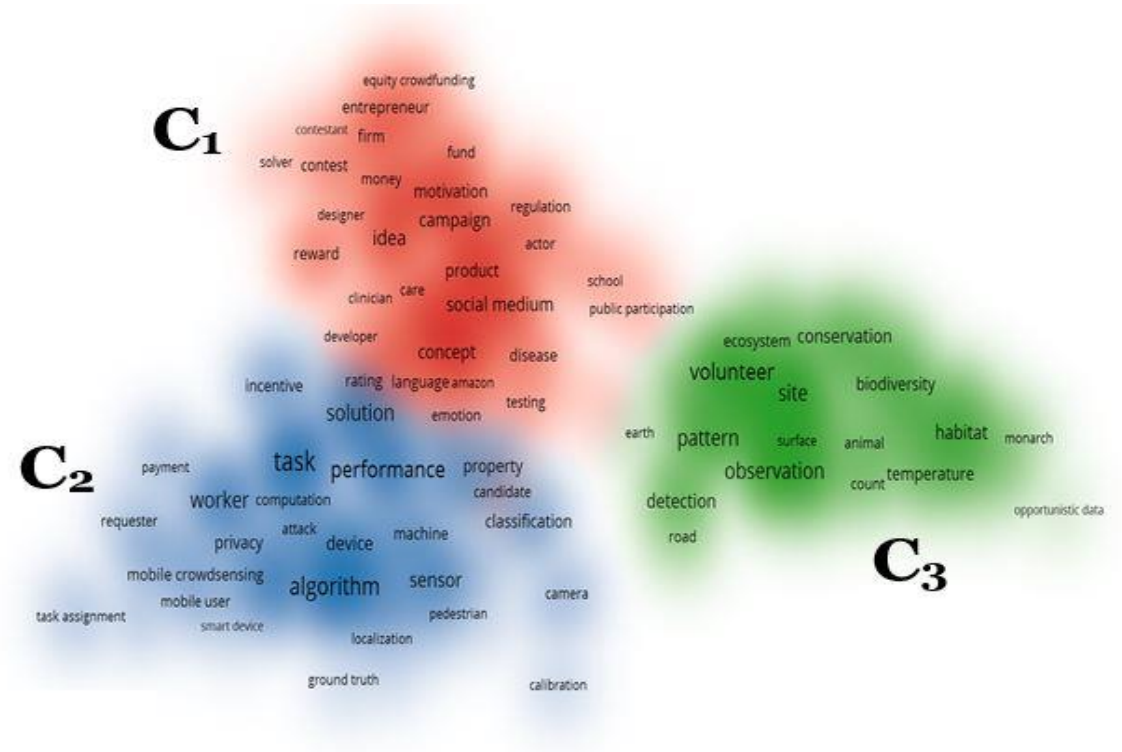

Fig 4: Cluster of occurring terms[1]

1. Cluster 1

Cluster 1 in Figure 4 has, at its core, *campaign, idea, and product*. Coupled with the 5 most frequent words for this cluster in Table III (*Idea, Concept, Campaign, Social medium, Product*), this cluster clearly points to the broader theme of ideation contests, whereby companies outsource their innovation activities to crowds via innovation campaigns through platforms [44]. We also spot an isolated-mini cluster within, symbolized with equity *crowdfunding, entrepreneur, and firm*. This mini-cluster points to start-up development through crowdfunding [45]. Thus, the broader theme for this idea emerges as “innovation”,

[1] “For coloured versions of all the clustering results, please see the online version.”

whether by innovating for big companies or funding for start-ups. Therefore, we label the theme in this cluster "Cluster 1: Crowdsourcing and Innovation (C1)".

2. Cluster 2

With words such as *algorithm, task, worker, performance,* and *solution*; Cluster 2 points to the general mechanism of crowdsourcing application systems [46]. This is further reinforced by words such as sensor, smart device, privacy, computation, mobile user, which indicate the general development of crowdsourcing applications [47]. This motivates the labelling of Cluster as "Cluster 2: Crowdsourcing and Engineering (C2)". We also want to emphasize how *solution* and *incentive* are co-opted by Cluster 1 and Cluster 2. This is because incentive mechanisms and solution-seeking are relevant concepts for both the general engineering applications and innovation campaigns.

3. Cluster 3

Cluster 3 seems to standalone with keywords such as *volunteer, site*, and *observation*. Supported with other frequent words such as *detection, habitat, biodiversity*, and *ecosystem*; closer analysis shows that this cluster comprises all citizen science activities carried out by the crowd willing to participate in scientific data collection by reporting their observations [48]. Thus, we appropriately label this cluster "Cluster 3: Crowdsourcing and Science (C3)".

As the crowdsourcing field presents such a wide and multidisciplinary group of studies and applications, further analyses are conducted to garner a deeper understanding of these clusters. We further break down the crowdsourcing field, focusing on each cluster individually.

D. Analysis of Research Clusters

This section examines in-depth the research clusters identified previously. To perform these analyses, each clusters' data is examined separately as explained in the methodology section.

1. Cluster 1: Crowdsourcing and Innovation

Cluster 1 is defined by the concentration and linkage of terms, which reveal the use of crowdsourcing as a model or capability to enable firms to access internal and external sources of labour, wisdom, creativity and funding through users, consumers and stakeholders. Terms such as an *idea, campaign, capital, consumer, contest, fund, product, motivation* and *reward* can be seen in Figure 5.

To assist the researcher in defining and presenting the scope of the cluster, the web of science field categories was utilised, this included: *business, economics, engineering industrial, public administration, law, hospitality leisure, sports tourism* etc. The prominent use of crowdsourcing applications ranges from *idea crowdsourcing, crowdfunding, crowd creation* as established previously on a broad scale. The sub-clusters can be dispersed into three main sub-clusters as A, B and C as shown in Figure 5 and these sub-clusters are further examined in the following sections.

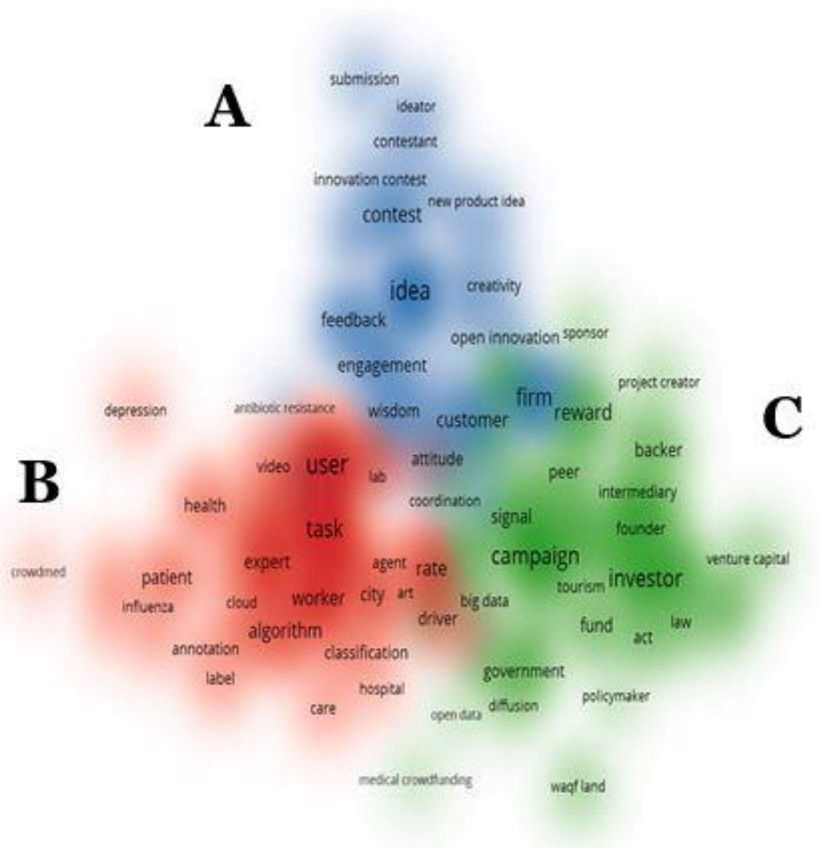

Fig 5: Clusters of Crowdsourcing and Innovation Keywords

1.1. Sub-cluster 1.A: Idea and Wisdom

Research in this sub-cluster focuses on the use of crowdsourcing for NPD and on tasks related to the engagement with the crowd within innovation activities (please see terms such as idea, contest and engagement). As organisations expand their boundaries and reach out to employees and external communities during stages of product development; areas of investigation arise related to idea quality, feedback, wisdom of the crowd, increasing new product market value, collective intelligence, customer ideation, identifying new product ideas [49 - 51]. Another area of focus in this cluster is the motivation and engagement of the crowd during competitions or contests with studies focused on idea competitions, idea implementation based on idea popularity, task design, participation in contests, recruiting valuable participants and modelling prizes [52 - 53].

1.2. Sub-cluster 1.B: Micro and Macro Tasks

Research in this sub-cluster investigates crowdsourcing as a tool for solving tasks, which can be broken down into studies such as crowdsourcing for HIV testing interventions, review of videos for research [54], optimal task allocations, improving consensus scoring, leveraging non-expert workers, identifying reliable workers [55, 56], and finally the use of crowdsourcing for mapping activities and disaster management [57]. The sub-cluster remains related to the main theme of innovation, given that task assignments are implemented in order to harvest more knowledge for a faster development process.

1.3. Sub-cluster 1.C: Donation and Investment

In line with the mini-cluster in the Cluster 1, the research focus here is on the funding of innovative projects, campaigns, and start-ups, with studies on problems hindering promised rewards, motivation to crowdfund and signalling in crowdfunding campaigns (please see terms such as entrepreneur, campaign, investor) [58, 59].

2. Cluster 2: Crowdsourcing and Engineering

As established previously, C2 focuses on the general mechanisms of crowdsourcing, identified by the existence of keywords such as *annotation, ground truth, map, device, incentive, mechanism, fingerprint,* and *sensor*.

The Web of Science categories utilised in this research theme are *computer science, information systems, telecommunications, chemistry analytical, engineering civil, transportation, engineering environmental,* and *regional urban planning, etc*. We observe that studies in C2 investigate the prominent applications of crowdsensing such as mobile crowdsourcing, spatial crowdsourcing, and volunteered geographic information; and tackle the general problems such as accuracy, trust, incentive, etc. Figure 6 illustrates the findings of the analysis.

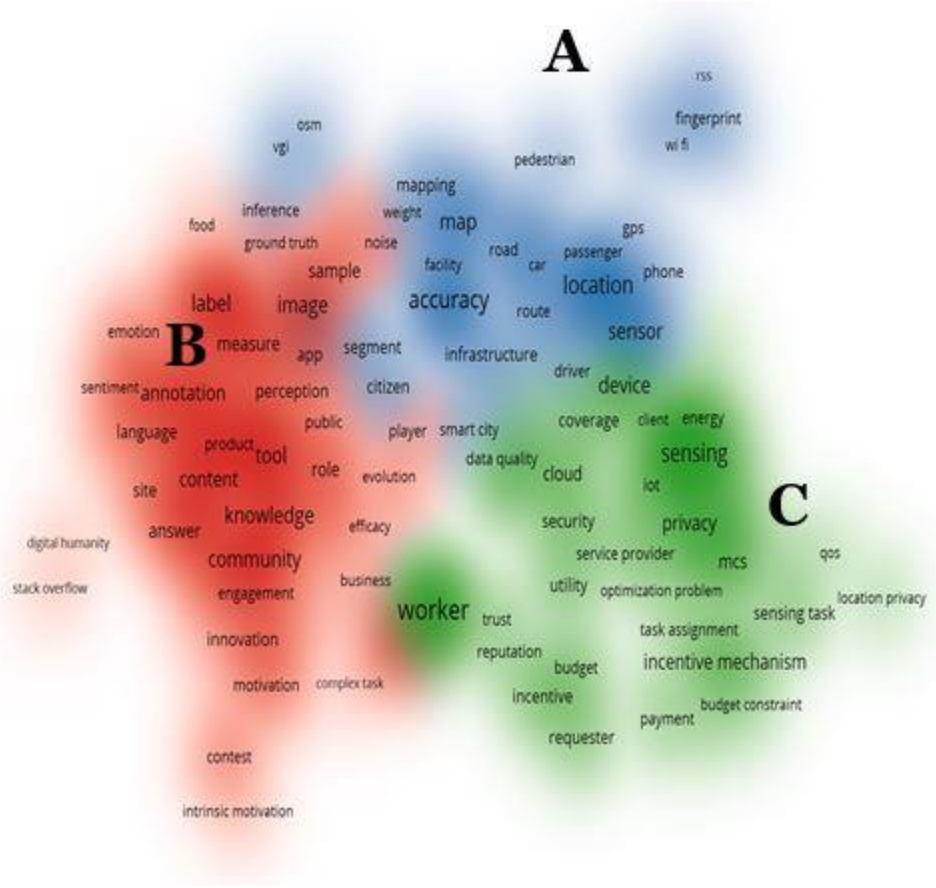

Fig 6: Clusters of Crowdsourcing and Engineering Keywords

2.1. Sub-cluster 2.A: Mapping

This sub-cluster is comprised of keywords such as a *map, location, GPS, route, etc*. These keywords point, first and foremost, to crowdsourcing tasks related to the improvement of geographic information systems and acquiring geographic information about the earth and environment, which can be disseminated via social media or collaborative projects such as Flickr, Twitter, Facebook and OpenStreetMap [60]. Another stream of studies in this sub-cluster focus on indoor localisations, path estimation, and floor plan construction [61, 62]. A third stream studies the use of devices such as smartphones and sensors to perform crowdsourcing activities related to mapping and construction, estimation of road conditions, and applications in smart city [63, 64].

2.2. Sub-cluster 2.B: Labour and Knowledge

This sub-cluster is illustrated with keywords such as *knowledge, community, annotation, label, content, motivation*; pointing to the clickworkers, their engagement, and their labour. These keywords describe two types of literature. First of all, we observe studies investigating tasks with relation to human assessments for facial image quality, rating images from photo-sharing websites, and language processing [65, 66]. The second stream of research in the sub-cluster is focused on the presence of an online community with a variety of skill set, benefits of human intelligence and the extraction of knowledge [67].

2.3. Sub-cluster 2.C: Architecture and Design

This sub-cluster implicates research related to the design of the crowdsourcing mechanism with keywords such as *task assignment, incentive mechanism, worker,* and *budget*. Studies here investigate design mechanisms for the assignment of tasks to the crowd and incentive schemes [68, 69] as well as privacy preservation scheme for the crowd whilst performing spatial crowdsourcing and mobile crowdsourcing tasks [70]

It is worth noting that Sub-clusters A and C overlap around *privacy* and *security*, explained by the existence of sensing and mobile crowdsourcing (MCS in Figure 6). These applications collect data from mobile devices, which creates privacy concerns which are investigated both in the 2.A and 2.C.

3. Cluster 3: Crowdsourcing and Science

This cluster contains many terms from natural sciences such as *amphibian, beach, butterfly, species, bird, egg, forest, galaxy, habitat, island, parasite, plant,* and *wildlife*. Coupled with keywords such as *camera, trap, conservation, planning, disaster, engagement,* and *image*, C3 implicates the popular application of crowdsourcing in natural sciences.

The Web of Science categories for this research theme includes *biodiversity, conservation, geography, environmental studies, water resources, oceanography, etc*. The prominent use of crowdsourcing applications in C3 is crowd science, citizen science, volunteer geographic information, participatory crowdsourcing, and passive crowdsourcing.

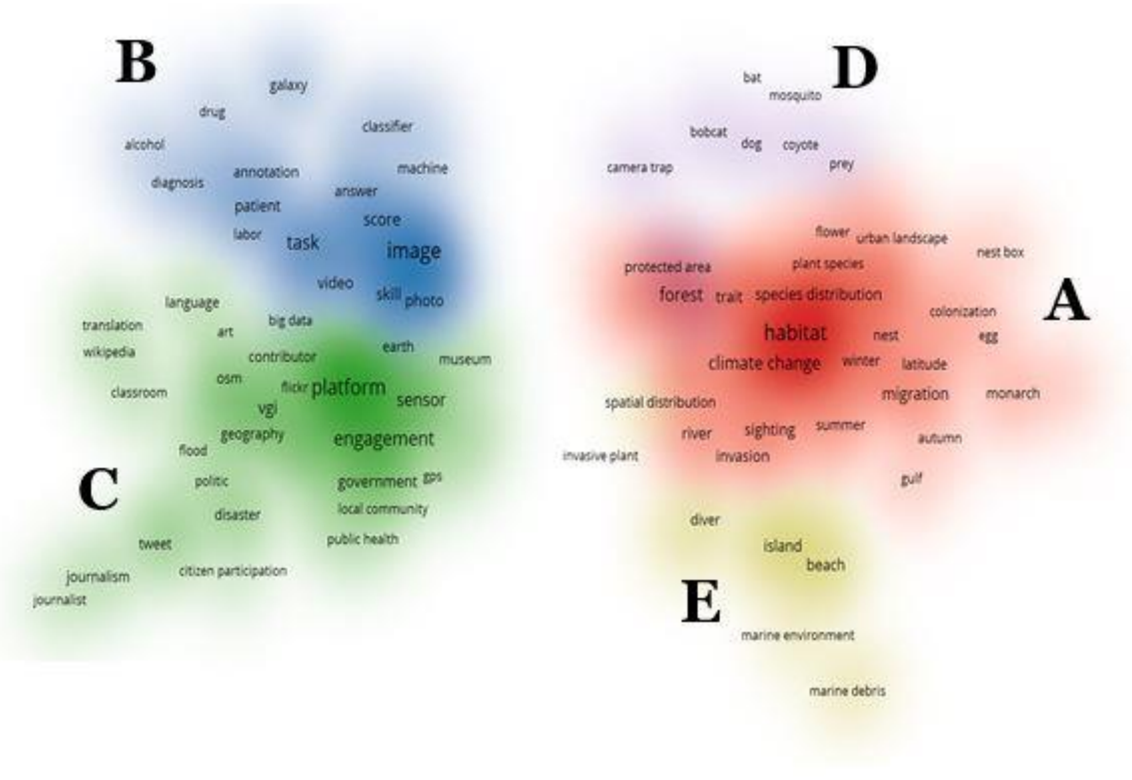

Fig 7: Clusters of Crowdsourcing and Science Keywords

To investigate temporal trends in more detail, we use the VOSviewer software which assists in the identification of closely related terms with further clustering of this research theme into sub-domains, illustrated in Figure 7. C3 displays a relatively dispersed cluster, with five sub-clusters emerging. Even though there is a lower semantic similarity of the keywords within C3, the investigation into the studies shaping the sub-clusters helps us identify the common themes for each sub-cluster.

3.1. Sub-cluster 3.A: Habitat Monitoring

Keywords such as *habitat, species distribution, migration,* and *sighting* are related to crowdsourcing tasks on monitoring nature by workers, ergo crowdsourcing for biology. Research in this area includes pollination by insects, the attractiveness of flowers to pollinators and impact of pesticides on insects [71, 72]; distribution of butterfly species during seasons, climate change impact on population trends on moth and butterfly species and temperature induced changes in plants [73].

3.2. Sub-cluster 3.B: Classification

Keywords such as *classification, image, accuracy* and *deep learning* point to crowdsourcing tasks related to general image annotation applications. Research in this sub-cluster includes quality assurance within health care and patient safety [74, 75] as well as machine learning for identification of bubbles, earth observation, enhancing image precision and image coverage [48, 76].

3.3. Sub-cluster 3.C: Public Engagement

Terms such as *disaster, flood, politic, tweet, museum, Wikipedia* and *classroom* point to varying streams of engagement in social tasks. Research here deals with spatial collective intelligence, humanitarian mapping, producing digital geospatial artefacts [77, 78], public participation in science-related projects that influence resource management and policies, public understanding of science, conservation outcomes, and engagement models [79].

3.4. Sub-cluster 3.D: Wildlife Preservation

Terms such as *bat, mosquito, bobcat* and *coyote* point to detections tasks for wildlife preservation. Research in this sub-cluster comprises of studies on the trends in bat populations, the influence of citizen science on conservation attitude and behaviours, urban ecosystem relationship between humans and coyote, differential responses of bat species, detection of invasive mosquitos [80 - 82].

3.5. Sub-cluster 3.E: Marine Conservation

Terms in this sub-cluster are similar to 3.D in terms of scarcity of keywords. A varying number of terms such as *island, beach, shark, marine debris* and *marine environment* are related to crowdsourcing tasks for marine conservation. Research in this sub-cluster comprises of studies on the distribution of small plastic debris on beaches, reproductive seasonality of fisheries, air temperature data collection, monitoring sea turtle populations [83 - 85].

V. DISCUSSION

Gathering all the investigations and findings from Section IV and the clustering results, we identify and classify the related clusters emerging within the crowdsourcing field as well as sub-clusters with related tasks involved to perform crowdsourcing activities. This is captured in Figure 8. Our findings attest to the growth of the field since the mid-2000s, which today can be categorized into three major research areas: Innovation, Engineering, and Science. These areas, even though not distinct, are nevertheless unique enough to be investigated separately. Our in-depth analysis showcases very different research streams stemming from these three areas.

Research in C1 targets innovation management journals such as Research Policy [52], Journal of Product Innovation Management [51], Technological Forecasting and Social Change [53], whereas studies in C2 are generally published in engineering journals such as multiple IEEE outlets [68 - 70]. Conversely, scholars opt for science journals such as Ecological Entomology [71], Biological Conservation [72], Current Surgery Reports [74], and Public Understanding of Science [79] for C3.

Comparing the results of Figure 8 to the relevant studies, we categorise crowdsourcing holistically without any limitations to certain research fields or to a number of studies, and our results identify many categories and sub-categories which are not categorised by other scholars. Comparing our study to Hossain and Kauranen [9], we also find the same areas such as idea generation, micro-tasking, and citizen science. However, we identify many other categories such as *donation* and funding and its relevant subcategories which are not apparent in their categorisation. Some of the categories that they discuss, such as open-source software and wikis, appear to be linked to the applications of the categories that we identified. Comparing our findings to Sivula and Kantola [7] who divided crowdsourcing into seven main categories, we identify all the same categories, but we contribute to the field by grouping them within innovation, engineering and scientific science clusters and also in their relevant subcategories. Hence, we organise crowdsourcing knowledge with a hierarchical and scientometric approach.

In the final step of our study, we map our findings from the literature to the real-life platforms and techniques, presented in Table IV. *Applications* present all the current applications and techniques emerging from our review: crowdwatch, crowd debugging, crowd science, civic crowdfunding, mobile crowdsourcing. We use our search string alongside the term "application" to assist in the identification of the typologies. There is a high prominence in the use of mobile crowdsourcing and crowd sensing approaches, showing the trend shift from just computer science and business domain to domains such as transportation, health, medicine, construction and social science fields. Research into the development crowdsourcing applications such as crowd science in terms of data quality, mobilizing participants involvement in fields such as social science would be a great avenue for research for utilizing this approach to certain tasks [86]. The last column demonstrates platforms and/or techniques using those specific applications.

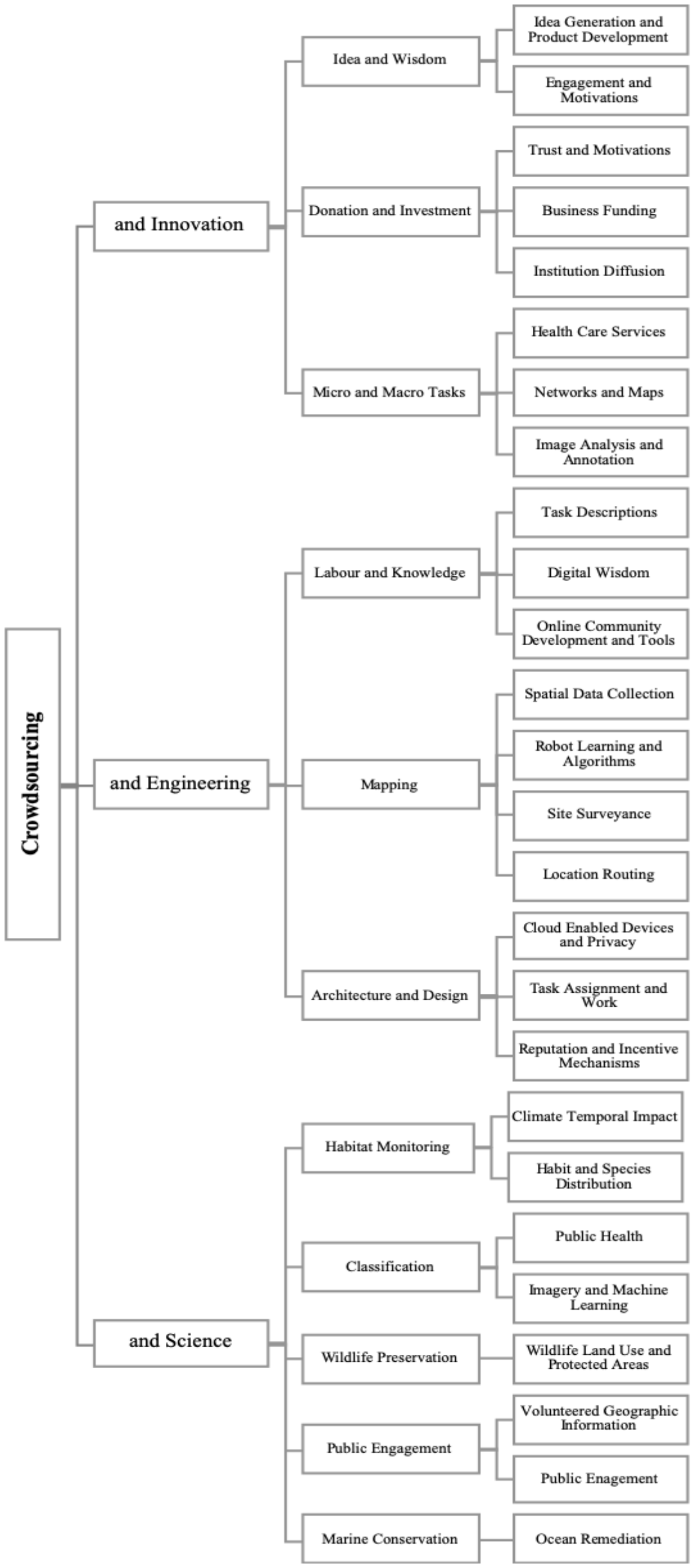

Fig 8: Breakdown of Crowdsourcing Domains

Table IV: Crowdsourcing Applications

Application	Application Typologies	Platforms/Techniques
Crowdsensing	Mobile Crowdsensing, Citizen sensing, Sparse Crowdsensing	CrowdTracker, MobiGroup, IONavi
Crowd Testing	Crowdbased testing, Software development, Software Crowdsourcing, Coding, Design, Qoe Crowdtesting	CrowdBuild, Topcoder, CrowdOracles, Crowd Debugging, Innovation Contest, Crowdsmelling, CrowdEV, Code Hunt, Stack Overflow, GitHub, Open Source Software, AppCheck.
Crowdfunding	Crowdfunded Journalism, Equity Crowdfunding, P2P Crowdfunding, Reward Based Crowdfunding, Civic crowdfunding	Kickstarter, Crowdcube, Syndicate room, Gofundme, Indiegogo, Seedrs, Patreon, Crowdfunder, RocketHub, LendingClub, Angelist, Prosper
Citizen science	Volunteered Geographic Information, Smartphone Citizen science, Volunteer Computing, Crowd science	Phylo, Safari Science, CoralWatch, Foldit, Season Spotter, CrowdCurio, SeaCleaner, Google Earth, Amazon Mechanical Turk
Micro Tasking/ Macro task Crowdsourcing	Cloud sourcing, Emergency Information Systems	Foodswitch, Fiverr, Amazon Mechanical Turk, ReCaptcha
Mobile Crowdsourcing	Mobile Crowd computing, Volunteered Geographic Information, Spatial Crowdsourcing, Mobile Crowdsensing	SmartSource, MobiCS, CrowdMonitor, CrowdPic, NoiseSense, Crowdsourcing Air Quality, CrowdSenSim, CrowdWIFI, MapLocal, Voice App, CityCare, Project Spear, Project Jagriti, AppLERT, Pazl, FlySensing, ShopProfiler, CrowdGIS, Clothes Radar, CRATER, Buy4Me, CrowdTracker, CrowdWatch, FindingNemo, FlierMeet, Hysense, WeCrowd, SecureFind, NoiseCo, CrowdOut, Txteagle
Crowd computing	Crowd social media computing	Wildlife@Home, Blockchain, CrowdEyes, SETI@home.
Crowd creation	Idea Crowdsourcing, Co-creation, Design Crowdsourcing	Ideastorm, Ideascale, Innocentive, Eyeka, Chaordix, Fiat, Muji, Lego, Jovoto
Crowdsourcing Systems	Open source software, Cloud Computing, Vehicular Fog Computing	OpenStreetMap, System Medicine, Crisis Mapping
Crowd Wisdom	Crowd sharing, Crowd networking, Fan sourcing, Crowd rating, Crowd voting	Youtube, Twitter, BzzAgent, Facebook, LinkedIn, Wikipedia, Delicious, Needle, Zuberance.

Table V: Future Research Directions for Crowdsourcing

Research Clusters	Sub-domains and Exemplary References	Future Research Avenues (Compiled from Literature)
Innovation	Idea and Wisdom [49 - 53]	Explore the impact of participation feedback on participants intensity during crowdsourcing ideation initiatives (consumer based innovation contests, internal or external tournaments); Further explore combination of participants and organization teams in arriving at quality ideas.
	Micro and Macro Tasks [54 - 56]	Further explore the role of campaign features in fostering participant loyalty and data quality; Best practises for merging trained experts and non-expert judgements for detection of improper tasks; Modeling patterns for crowd task completion.
	Donation and Investment [58 - 59]	Explore crucial signals for business ventures in reward based crowdfunding success; Explore best mechanisms managing fraud and reward delivery delays in reward-based crowdfunding.
Engineering	Labour and Knowledge [65 - 67]	Understand the curation of knowledge by comparing labour and knowledge communities.
	Mapping [60], [62]	Explore the implementation of crowdsourced Wi-Fi fingerprints to building-scale spaces and open graph areas; Investigate power consumption and security conscious crowdsourcing of Wi-Fi fingerprints.
	Architecture and Design [68 - 70]	Explore architectural authentication methods to achieve evictions and concealing of misbehaving workers identity; Best practises or factors making smartphone applications more appealing for crowd task performance; Best practices for designing incentive mechanisms that preserve participant's privacy.
Science	Habitat Monitoring [71], [73]	Best practices for recruiting experienced participants for crowdsourcing monitoring tasks; Explore factors and varying participant features affecting data quality in citizen science initiatives.
	Classification [48], [74], [76]	Explore the combination of crowdsourcing and deep learning techniques for wider classification tasks (e.g national and regional human settlement data); Best practices to combine human and machine searches in scaling large image data sets.
	Wildlife Preservation [80], [81], [82]	Contextualizing relations between motivations, project participation and resulting outcomes; Attitudes towards wildlife for mitigation of human-wildlife conflicts.
	Public Engagement [77], [78], [79]	Explore the breakdown of virtual communities such as OpenStreetMappers from social sciences perspective (e.g nationality, demography or socio-economic status); Explore the effect of trust, fairness, equity and risk in volunteer recruitment, protocols, and data quality.
	Marine Conservation [83 - 85]	Explore the combination of port surveys and virtual communities in monitoring rare and threatened species; Best practises for developing mobile application for marine species data collection.

Based on the sub-domains of crowdsourcing as illustrated in Figure 8, the future research avenues for crowdsourcing is provided in detail as shown in Table V. The listed future research avenues are written based on key references in light of the previously completed analysis and also holistic view of the clustering results.

VI. CONCLUSIONS

The motivation of this study stemmed from the lack of a review of the crowdsourcing concept from a holistic perspective. Our findings contribute to the stream of literature on crowdsourcing by providing a scientometric-based methodological analysis of its use in the domains of science, engineering and innovation [1, 7-10]. Significantly we identify new areas for research. This study presents the analysis of data relating to publications advancing the field in crowdsourcing from 2006 to 2019, thereby offering emerging research themes and sub-clusters to researchers, experts, and the crowdsourcing community, along with implications to companies, managers, and practitioners. A total of 7,059 scientific publications have been identified during this period with an exponential increase in computer science (26%), environmental sciences (15%), engineering (14%), and business (10%). The keyword analysis of publications further reveals a concentration of research within three main emerging clusters with a range of top trending terms within each cluster. Examination of these main clusters reveal sub-clusters in relation to task and research: crowdsourcing and innovation (i. idea and wisdom ii. micro and macro tasks iii. donation and investment), crowdsourcing and engineering (i. mapping ii. labour and knowledge iii. architecture and design), crowdsourcing and science (i. habitat monitoring ii. classification iii. public engagement iv. wildlife preservation v. marine conservation). Whilst examining the trend of crowdsourcing applications, we discovered applications such as crowdsensing, crowdtesting, crowdfunding, citizen science, micro-tasking/macro-tasking, mobile crowdsourcing, crowd computing, crowd creation, and crowd wisdom.

One of the practical implication of this paper is the mapping crowdsourcing research and applications holistically considering innovation, engineering and science domains. The clustering, categorisation and sub-categorisation results are further linked with the relevant applications and hence it provides a hierarchical taxonomy for other scholars and industrial practitioners. We illustrate the results with a number of examples to show a broad spectrum of crowdsourcing applications and methods in different conditions. We clarify the linkage between each research cluster and sub-clusters are examined to show the interrelationship of crowdsourcing research. We also illustrate emerging or developing fields for relevant practitioners to take strategic actions.

The methodological contributions of this paper are threefold. Firstly, crowdsourcing specific search string is developed so other scholars can perform similar studies by using our search string. Secondly, we provide a highly transparent process of how the reliability and validity of such a study can be increased in terms of data retrieval. Finally, we provide a great example to granularity in such methods where clustering is linked to the sub-clusters, and subclusters are linked to its relevant categories and applications.

The study, however, is not without limitations. We adopted a scientometric approach combining a co-occurrence text mining and publication analysis to review the literature. We have used a range of keywords in our search within the abstract, title and full text of publications. An expansion of keywords utilised may generate different search results. We have classified crowdsourcing applications into three categories and this classification is by no means exhaustive, thereby requires other studies to consolidate the findings. We have provided useful insights into the growth and development of the crowdsourcing field.

Acknowledgement

The contributions in this article by Dr Sercan Ozcan was prepared within the framework of the Basic Research Program of the National Research University Higher School of Economics.

REFERENCES

[1] Y. Zhao and Q. Zhu, "Evaluation on crowdsourcing research : Current status and future direction," *Information Systems Frontiers*, vol. 16, no. 3, July, 2014, pp. 417–434.

[2] M. W. G. Rocha, A. S. O. Yu, and P. T. de Souza Nascimento, "Crowdsourcing in the Fuzzy Front End of innovation," *Manag. Eng. Technol. (PICMET), Portl. Int. Conf.*, IEEE, July, 2014, pp. 830–839.

[3] B. J. Howe, "The Rise of Crowdsourcing," *Wired magazine*, vol. 6, no.14, June, 2012, pp. 1–5.

[4] D. C. Brabham, "Crowdsourcing as a Model for Problem Solving An Introduction and Cases," *Convergence*, vol. 14, no. 1, Feb, 2008, pp. 75–90.

[5] M. Hosseini, A. Shahri, K. Phalp, and R. Ali, "Recommendations on adapting crowdsourcing to problem types," *Proc. - Int. Conf. Res. Challenges Inf. Sci.*, IEEE, May, 2015, pp. 423–433.

[6] I. Blohm, J. M. Leimeister and H. Krcmar, "Crowdsourcing: How to Benefit from (Too) Many Great Ideas," *MIS Quarterly Executive*, vol. 12, no. 4, 2013, pp. 199–211.

[7] A. Sivula and J. Kantola, "Integrating crowdsourcing with holistic innovation management," *International Journal of Advanced Logistics*, vol. 5, no. 3-4, Nov, 2016, pp. 153-64.

[8] M. Palacios, A. Martinez-Corral, A. Nisar, and M. Grijalvo, "Crowdsourcing and organizational forms: Emerging trends and research implications," *Journal of Business Research*, vol. 69, no. 5, May, 2016, pp. 1834–1839.

[9] M. Hossain and I. Kauranen, "Crowdsourcing: a comprehensive literature review," *Strateg. Outsourcing An Int. J.*, vol. 8, no. 1, Feb, 2015, pp. 2–22.

[10] Z. Zakariah, N. Janom, N. H. Arshad, S. S. Salleh, and S. R. S. Aris, "Crowdsourcing: The Trend of Prior Studies," *Proc. - 2014 4th Int. Conf. Artif. Intell. with Appl. Eng. Technol. ICAIET 2014, IEEE*, Dec, 2014, pp. 129–133.

[11] N. Islam and S. Ozcan, "The management of nanotechnology: analysis of technology linkages and the regional nanotechnology competencies," *R&D Management*, vol. 47, no. 1, Jan, 2017, pp. 111–126.

[12] I. Rafols, M. M. Hopkins, J. Hoekman, J. Siepel, A. O'Hare, A. Perianes-Rodriguez, P. Nightingale, "Big Pharma, little science?. A bibliometric perspective on Big Pharma's R&D decline," *Technol. Forecast. Soc. Change*, vol. 81, no. 1, Jan, 2014, pp. 22–38.

[13] A. Kovács, B. Van Looy, and B. Cassiman, "Exploring the scope of open innovation: a bibliometric review of a decade of research," *Scientometrics*, vol. 104, no. 3, Sep, 2015, pp. 951–983.

[14] M. Hosseini, A. Shahri, K. Phalp, J. Taylor, and R. Ali, "Crowdsourcing : A taxonomy and systematic mapping study," *Computer Science Review*, vol. 17, Aug, 2015, pp. 43–69.

[15] A. C. Rouse, "A Preliminary Taxonomy of Crowdsourcing," *ACIS 2010*

Proceedings, vol. *76*, Dec, 2010, pp. 1-10.
[16] E. Schenk and C. Guittard, "Crowdsourcing : What can be Outsourced to the Crowd , and Why," In *Workshop on open source innovation, Strasbourg, France*, vol. 72, Dec, 2012, p. 3.
[17] E. Estellés-Arolas and F. González-Ladrón-de-Guevara, "Towards an integrated crowdsourcing definition," *Journal of Information science*, vol. 38, no. 2, Apr, 2012, pp. 189–200.
[18] L. Hetmank, "Components and Functions of Crowdsourcing Systems – A Systematic Literature Review," *Wirtschaftsinformatik,* no. 4, March, 2013, pp. 55–69.
[19] R. T. Nakatsu, E. B. Grossman, and C. L. Iacovou, "A taxonomy of crowdsourcing based on task complexity," *Journal of Information Science*, vol. *40, no.* 6, Dec, 2014, pp. 823-834.
[20] B. M. Good and A. I. Su, "Genome analysis Crowdsourcing for bioinformatics," *Bioinformatics,* vol. 29, no. 16, June, 2013, pp. 1925–1933.
[21] B. Solemon, I. Ariffin, M. M. Din, and R. M. Anwar, "A review of the uses of crowdsourcing in higher education," *International Journal of Asian Social Science*, vol. 3, no. 9, 2013, pp. 2066–2073.
[22] J. Pedersen, D. Kocsis, A. Tripathi, A. Weerakoon, N. Tahmasbi, G. J, De Vreede and A. Tarrell, "Conceptual Foundations of Crowdsourcing : A Review of IS Research," In *2013 46th Hawaii International Conference on System Sciences IEEE*, Jan, 2013, pp. 579 - 588.
[23] D. C. Brabham, K. M. Ribisl, T. R. Kirchner, and J. M. Bernhardt, "Crowdsourcing Applications for Public Health," *Am. J. Prev. Med.*, vol. 46, no. 2, pp. 179–187, Feb, 2014.
[24] B. L. Ranard, Y.P Ha, Z. F. Meisel, D. A. Asch, S. S. Hill, L.B. Becker and R. M. Merchant, "Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review," *Journal of general internal medicine*, vol. 29, no. 1, pp. 187–203, Jan, 2014.
[25] E. Estellés-arolas, I. K. Hansson, T. Ajtamurto, T. Ludwig and M. Muller, "A task-based crowdsourcing typology," *Guest Editors*, no 57. Dec, 2016.
[26] R. Buettner, "A systematic literature review of crowdsourcing research from a human resource management perspective," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, IEEE, Jan, 2015, pp. 4609–4618,.
[27] J. Minet, Y. Curnel, A. Gobin, J. P. Goffart, F. Melard, B. Tychon, and P. Defourny, "Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach," *Computers and Electronics in Agriculture*, vol. *142*, Nov, 2017, pp. 126-138.
[28] T. Ambreen and N. Ikram, "A state-of-the-art of Empirical Literature of Crowdsourcing in Computing," In *2016 IEEE 11th International Conference on Global Software Engineering (ICGSE),* Aug, 2016, *pp*. 189 - 190.
[29] A. Ghezzi, D. Gabelloni, A. Martini, and A. Natalicchio, "Crowdsourcing: A Review and Suggestions for Future Research," I*nternational Journal of Management Reviews*, vol. 20, no. 2, Apr, 2018, pp. 343 - 363.
[30] O. Zuchowski, O. Posegga, D. Schlagwein, and K. Fischbach, "Internal crowdsourcing: Conceptual framework, structured review, and research agenda," *Journal of Information Technology*, vol. 31, no. 2, Jun, 2016, pp. 166–184.
[31] A. Tripathi, N. Tahmasbi, D. Khazanchi, and L. Najjar, "Crowdsourcing Typology : A Review of IS Research and Organizations," *Proceedings of the Midwest Association for Information Systems (MWAIS)*, 2014.

[32] B. A. Malik, A. Aftab, and N. A. Pm, “Mapping of Crowdsourcing Research : A Bibliometric Analysis,” *Journal of Library & Information Technology DESIDOC,* vol. 39, no. 1, Jan, 2019, pp. 23–30.
[33] M. Ertz and S. Leblanc-Proulx, “Sustainability in the collaborative economy : A bibliometric analysis reveals emerging interest,” *Journal of Cleaner Production*, vol. 196, Sep, 2018, pp. 1073–1085.
[34] A. Darko, A. P. Chan, X. Huo and D. G. Owusu-Manu, “A scientometric analysis and visualization of global green building research,” *Building and Environment*, vol. 149, Feb, 2019, pp. 501–511.
[35] H. N. Su and P. C. Lee, “Framing the structure of global open innovation research,” *Journal of Informetrics*, vol. *6, no.* 2, April, 2012, pp. 202-216.
[36] D. Chatterjee and S. Sahasranamam, “Trends in innovation management research in India–an analysis of publications for the period 1991–2013,” *Current Science*, vol. *107, no.* 11, 2014, pp. 1800-1805.
[37] F. P. Appio, A. Martini, S. Massa and S. Testa, “Unveiling the intellectual origins of Social Media-based innovation: insights from a bibliometric approach,” *Scientometrics*, vol. *108, no.* 1, 2016, pp. 355-388.
[38] C. Kullenberg and D. Kasperowski, “What Is Citizen Science ? – A Scientometric meta-analysis,” *PloS one*, vol. 11, no. 1, Jan, 2016, pp. 1–16.
[39] L. Mora, M. Deakin, and A. Reid, “Combining co-citation clustering and text-based analysis to reveal the main development paths of smart cities.” *Technological Forecasting and Social Change*, vol. *142*, May, 2019, pp. 56-69.
[40] Y. Yu, H. Wang, G. Yin, and T. Wang, “Reviewer Recommendation for Pull-Requests in GitHub : What Can We Learn from Code Review and Bug Assignment ?,” *Information and Software Technology*, vol. 74, Jun, 2016, pp. 204-218.
[41] N. Silva, E. R. Marques and L. Lopes, “Flux : A Platform for Dynamically Reconfigurable Mobile Crowd-sensing,” *ACM Transactions on Sensor Networks (TOSN)*, vol. *14, no.* 3-4, Nov, 2018, pp. 20.
[42] G. Geisler, G. Willard and C. Ovalle, “A crowdsourcing framework for the production and use of film and television data,” *New Review of Hypermedia and Multimedia,* vol. 17, no. 1, Apr, 2011, pp. 73-97.
[43] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang and X. Zhou, “Mobile crowd sensing and computing : The review of an emerging human-powered sensing paradigm,” *ACM Computing Surveys (CSUR),* vol. 48, no. 1, Sep, 2015, pp. 7.
[44] Y. Huang, P. Vir Singh, and K. Srinivasan, “Crowdsourcing New Product Ideas Under Consumer Learning,” *Management Science*, vol. 60, no. 9, Jul, 2014, pp. 2138–2159.
[45] E. Mollick, “The dynamics of crowdfunding : An exploratory study,” *Journal of Business Venturing*, vol. 29, no. 1, Jan, 2014, pp. 1–16.
[46] W. Mason and S. Suri, “Conducting behavioral research on Amazon’s Mechanical Turk,” *Behavior research methods*, vol. 44, no. 1, Mar, 2012, pp. 1–23.
[47] C. Wu, Y. Liu and Z. Yang, “Smartphones Based Crowdsourcing for Indoor Localization,” *IEEE Trans. Mob. Comput.*, vol. 14, no. 2, April, 2015, pp. 444–457.
[48] C. N. Beaumont, A. A. Goodman, S. Kendrew, and J. P. Williams, “The milky way project: leveraging citizen science and machine learning to detect interstellar bubbles.”*The Astrophysical Journal Supplement Series*, vol. *214, no.* 1, Aug, 2014, pp. 3.
[49] M. Garcia and B. Walton, “The wisdom of crowds : The potential of online communities as a tool for data analysis,” *Technovation*, vol. 34, no. 4, Apr, 2014, pp. 203–214.

[50] D. Zahay, N. Hajli, and D. Sihi, “Managerial perspectives on crowdsourcing in the new product development process,” *Industrial Marketing Management*, vol. 71, May, 2018, pp. 41–53.

[51] M. K. Poetz and M. Schreier, “The value of crowdsourcing: Can users really compete with professionals in generating new product ideas?,” *Journal of Product Innovation Management*, vol. 29, no. 2, March, 2012, pp. 245–256.

[52] B. Schemmann, A. M. Herrmann, M. M. H. Chappin, and G. J. Heimeriks, “Crowdsourcing ideas: Involving ordinary users in the ideation phase of new product development,” *Research Policy*, vol. 45, no. 6, July, 2016, pp. 1145–1154.

[53] L. Mortara, S. J. Ford, and M. Jaeger, “Idea Competitions under scrutiny : Acquisition, intelligence or public relations mechanism ?,” *Technol. Forecast. Soc. Chang.*, vol. 80, no. 8, Oct, 2013, pp. 1563–1578.

[54] W. Tang, L. Han, J. Best, Y. Zhang, K. Mollan, J. Kim and S. Galler, “Crowdsourcing HIV Test Promotion Videos: A Noninferiority Randomized Controlled Trial in China,” *Clinical infectious diseases*, vol. 62, no. 11, Apr, 2016, pp. 1436–1442.

[55] A. Tarasov, S. J. Delany, and B. Mac Namee, “Dynamic estimation of worker reliability in crowdsourcing for regression tasks : Making it work,” *Expert Systems with Applications*, vol. 41, no. 14, Oct, 2014, pp. 6190–6210.

[56] Y. Baba, H. Kashima, K. Kinoshita, G. Yamaguchi, and Y. Akiyoshi, “Leveraging non-expert crowdsourcing workers for improper task detection in crowdsourcing marketplaces” *Expert Systems with Applications*, vol. 41, no. 6, May, 2014, pp. 2678–2687.

[57] C. Granell and F. O. Ostermann, “Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster management,” *Computers , Environment and Urban Systems* (*CEUS)*, vol. 59, Sep, 2016, pp. 231–243.

[58] M. M. Kunz, U. Bretschneider, M. Erler, and J. M. Leimeister, “An empirical investigation of signaling in reward-based crowdfunding,” *Electronic Commerce Research*, vol. 17, no. 3, Sep, 2017, pp. 425- 461.

[59] F. Schiavone, “Incompetence and Managerial Problems Delaying Reward Delivery in Crowdfunding,” *Journal of Innovation Economics Management*, no. 2, 2017, pp. 185-207.

[60] P. Mooney, P. Corcoran and B. Ciepluch, “The potential for using volunteered geographic information in pervasive health computing applications,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, no. 6, Dec, 2013, pp. 731–745.

[61] B. Zhou, Q. Li, W. Tu, X. Zhang L. Chen and Q. Mao, “ALIMC : Activity Landmark-Based Indoor Mapping via Crowdsourcing,” *IEEE Transactions on Intelligent Transportation Systems*, vol. *16, no.* 5, May, 2015, pp. 2774-2785.

[62] X. Zhang, A. K. Wong, C. Lea and R. S. K. Cheng, “Unambiguous Association of Crowd-Sourced Radio Maps to Floor Plans for Indoor Localization,” *IEEE Trans. Mob. Comput.*, vol. 17, no. 2, Jul, 2017, pp. 488–502.

[63] F. Kalim, J. P. Jeong and M. U. Ilyas, “CRATER : A Crowd Sensing Application to Estimate Road Conditions,” *IEEE Access*, vol. 4, Sep, 2016, pp. 8317–8326.

[64] H. Liu, Y. Zhou, and Y. Zhang, “Estimating Users’ Home and Work Locations Leveraging Large-Scale Crowd-Sourced Smartphone Data,” *IEEE Commun. Mag.*, vol. 53, no. 3, March, 2015, pp. 71–79.

[65] H. Zhai, T. Lingren, L. Deleger, Q. Li, M. Kaiser, L. Stoutenborough and I. Solti, “Web 2 . 0-Based Crowdsourcing for High-Quality Gold Standard Development in Clinical Natural Language Processing,”*Journal of medical Internet research*, vol. 15, no. 4, 2013, pp. 1–17.

[66] E. Siahaan, A. Hanjalic, and J. Redi, "A reliable methodology to collect Ground truth data of image aesthetic appeal," *IEEE Transactions on Multimedia*, vol. 18, no. 7, Apr, 2016, pp. 1338–1350.

[67] Y. B. Diven, "Communities of Knowledge : Creating and Connecting Resource Metadata," *Serials Review*, vol. 39, no. 3, Sep, 2014, pp. 162–166.

[68] Y. Gao, Y. Chen and K. R. Liu, "On Cost-Effective Incentive Mechanisms in Microtask Crowdsourcing," *IEEE Trans. Comput. Intell. AI Games*, vol. 7, no. 1, Jan, 2014, pp. 3–15.

[69] P. Cheng, X. Lian, L. Chen, J. Han, and J. Zhao, "Task Assignment on Multi-Skill Oriented Spatial Crowdsourcing," *IEEE Transactions on Knowledge and Data Engineering,* vol. 28, no. 8, Apr, 2016, pp. 2201 – 2215.

[70] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Security, privacy and incentive provision for mobile crowd sensing systems," *IEEE Internet Things J.*, vol. 3, no. 5, Apr, 2016, pp. 839–853.

[71] L. Birkin and D. Goulson, "Using citizen science to monitor pollination services," *Ecological Entomology*, vol. 40, Sep, 2015, pp. 3–11.

[72] A. Muratet and B. Fontaine, "Contrasting impacts of pesticides on butterflies and bumblebees in private gardens in France," *Biological Conservation*, vol. Feb, 2015, 148–154.

[73] B. Karlsson, "Extended season for northern butterflies," *International journal of biometeorology*, vol. 58, no. 5, Jul, 2014, pp. 691–701.

[74] S. Shackelford and M. Bowyer, "Modern Metrics for Evaluating Surgical Technical Skills," *Current Surgery Reports*, vol. 5, no. 10, Oct, 2017, pp. 1–10.

[75] L. T. Car, M. El-Khatib, R. Perneczky, N. Papchristou, R. Atun, I. Rudan, and A. Majeed, "Prioritizing problems in and solutions to homecare safety of people with dementia : supporting carers , streamlining care," *BMC Geriatr.*, vol. 17, no. Dec, 2017, 1–8.

[76] J. Frank, U. Rebbapragada, J. Bialas, T. Oommen, and T. C. Havens, "Effect of Label Noise on the Machine-Learned Classification of Earthquake Damage," Remote Sensing, vol. 9, no. 8, Aug, 2017, pp. 108.

[77] A. Ballatore and P. Mooney, "Conceptualising the geographic world: the dimensions of negotiation in crowdsourced cartography," *International Journal of Geographical Information Science*, vol. 29, no. 12, Dec, 2015, pp. 2310–2327.

[78] E. Vaz and J. J. Arsanjani, "Crowdsourced mapping of land use in urban dense environments : An assessment of Toronto," *The Canadian Geographer/Le Géographe canadien,* vol. 59, no. 2, June, 2015, pp. 246–255.

[79] R. Bonney, T. B. Phillips, H. L. Ballard, and J. W. Enck, "Can citizen science enhance public understanding of science ?," *Public Understanding of Science*, vol. *25, no.* 1, Jan, 2016, pp. 2-16.

[80] K. E. Barlow, P. A. Briggs, K. A. Haysom, A. M. Hutson, N. L. Lechiara, P. A. Racey and S. D. Langton "Citizen science reveals trends in bat populations : The National Bat Monitoring Programme in Great Britain," *Biol. Conserv.*, vol. 182, 2015, pp. 14–26.

[81] A. H. Toomey and M. C. Domroese, "Can citizen science lead to positive conservation attitudes and behaviors ?," vol. 20, no. 1, Jun, 2013, pp. 50-62.

[82] M. Weckel, D. A. Bogan, R. L. Burke, C. Nagy, and T. Green, "Coyotes Go ' Bridge and Tunnel ': A Narrow Opportunity to Study the Socio-ecological Impacts of Coyote Range Expansion on Long Island , NY Pre- and Post-Arrival," *Cities and the Environment (CATE),* vol. 8, no. 1, 2015, pp. 5.

[83] V. Hidalgo-ruz and M. Thiel, “Distribution and abundance of small plastic debris on beaches in the SE Pacific (Chile): A study supported by a citizen science project,” *Marine environmental research*, vol. 87, Jun, 2013, pp. 12–18.

[84] A. D. Syakti, “Microplastics Monitoring in Marine Environment,” *Omni-Akuatika* vol. 13, no. 2, Nov, 2017.

[85] J. L. Williams, S. J. Pierce, M. M. Fuentes, and M. Hamann, “Effectiveness of recreational divers for monitoring sea turtle populations,” *Endangered Species Research,* vol. 26, no. 3, Jan, 2015, pp. 209–219.

[86] R. Heiss and J. Matthes, “Citizen Science in the Social Sciences: A Call for More Evidence,” *GAIA-Ecological Perspectives for Science and Society*, vol. *26, no.* 1, Jan, 2017, pp. 22–26.

[87] L. Leydesdorff, “The challenge of scientometrics: The development, measurement, and self-organization of scientific communications,” Universal-Publishers, 2001.

[88] A. L. Porter and S. W. Cunningham, “Tech mining: Exploiting new technologies for competitive advantage,” John Wiley & Sons. vol. 29, no. 1, Jan 8, 2004, pp. 30-36.

[89] G. Marzi, F. Ciampi, D. Dalli, and M. Dabic, “New product development during the last ten years: the ongoing debate and future avenues,” *IEEE Transactions on Engineering Management*, Jun 16, 2020.

[90] S. Ozcan and A. Corum, “A Scientometric Analysis of Remanufacturing by Mapping Scientific, Organizational, and National Concentration Zones,” *IEEE Transactions on Engineering Management*, Jul 24, 2019.